Math 408 - Mathematical Statistics

Lecture 3. Discrete Random Variables

January 23, 2013

Agenda

- Random Variable: Motivation and Definition
- Cumulative Distribution Functions
- Properties of CDFs
- Discrete Random Variables
- Important Examples
  - The Point Mass Distribution
  - The Discrete Uniform Distribution
  - The Bernoulli Distribution
  - The Binomial Distribution
  - The Geometric Distribution
  - The Poisson Distribution
- Summary

Motivation and Definition

Statistics is concerned with data.

Question: How do we link sample spaces and events to data?

Answer: The link is provided by the concept of a **random variable**.

Definition

A random variable is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real number $x = X(\omega)$ to each realization $\omega \in \Omega$.

Remark: Technically, a random variable must be a measurable function.

Example: Flip a coin 10 times. Let $X(\omega)$ be the number of heads in the sequence. For example, if $\omega = HHTHTTTHTH$, then $X(\omega) = 5$.

Given a random variable X and a set $A \subset \mathbb{R}$, define

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

and let

$$\boxed{\begin{aligned} \mathbb{P}(X \in A) =& \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \\ \mathbb{P}(X = x) =& \mathbb{P}(X^{-1}(x)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \end{aligned}}$$

The Cumulative Distribution Function

Definition

The cumulative distribution function (CDF) $F_X : \mathbb{R} \to [0, 1]$ is defined by

$$\boxed{F_X(x) = \mathbb{P}(X \leq x)}$$

Example: Flip a fair coin twice and let X be the number of heads.
Find the CDF of X

Question: Why do we bother to define CDF?

Answer: CDF effectively contains all the information about the random variable

Theorem

Let X have CDF F and Y have CDF G. If $F(x) = G(x)$ for all x, then $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$. In words, the CDF completely determines the distribution of a random variable.

Properties of CDFs

Question: Given a function $F(x)$, can we find a random variable X such that $F(x)$ is the CDF of X, $F_X(x) = F(x)$?

Theorem

A function $F : \mathbb{R} \to [0, 1]$ is a CDF for some random variable if and only if it satisfies the following three conditions:

1. *F is non-decreasing:*

$$x_1 < x_2 \quad \Rightarrow \quad F(x_1) \leq F(x_2)$$

2. *F is normalized:*

$$\lim_{x \to -\infty} F(x) = 0 \quad \textit{and} \quad \lim_{x \to +\infty} F(x) = 1$$

3. *F is right-continuous:*

$$\lim_{y \to x+0} F(y) = F(x)$$

Discrete Random Variables

Definition

X is **discrete** if it takes countable many values $\{x_1, x_2, \dots\}$.
We define the **probability mass function** (PMF) for X by

$$\boxed{f_X(x) = \mathbb{P}(X = x)}$$

Example: Flip a fair coin twice and let X be the number of heads. Find the probability mass function of X.

The CDF of X is related to the PMF f_X by

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

The PMF f_X is related to the CDF F_X by

$$f_X(x) = F_X(x) - F_X(x^-) = F_X(x) - \lim_{y \to x-0} F(y)$$

Important Examples

- **The Point Mass Distribution**
  X has a point mass distribution at a, denoted $X \sim \delta_a$, if $\mathbb{P}(X = a) = 1$.
  In this case

$$F(x) = \begin{cases} 0, & x < a; \\ 1, & x \geq a. \end{cases}$$

  and

$$f(x) = \begin{cases} 1, & x = a; \\ 0, & x \neq a. \end{cases}$$

- **The Discrete Uniform Distribution**
  Let $n > 1$ be a given integer. Suppose that X has probability mass function given by

$$f(x) = \begin{cases} 1/n, & \text{for } x = 1, \ldots, n; \\ 0, & \text{otherwise.} \end{cases}$$

  We say that X has a uniform distribution on $1, \ldots, n$.

Important Examples

- **The Bernoulli Distribution**
  Let X represents a coin flip. Then $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ for some $p \in [0, 1]$. We say that X has a Bernoulli distribution, denoted $X \sim \mathrm{Bernoulli}(p)$. The probability mass function is

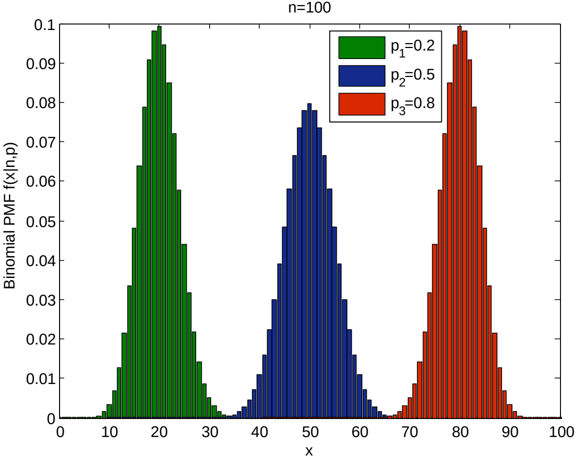
$$f(x|p) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}$$

- **The Binomial Distribution**
  Suppose we have a coin which falls heads with probability p for some $p \in [0, 1]$. Flip the coin n times and let X be the number of heads. Assume that the tosses are independent. The probability mass function of X is then

$$f(x|n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, \ldots, n; \\ 0, & \text{otherwise.} \end{cases}$$

  A random variable with this mass function is called a Binomial random variable and we write $X \sim \mathrm{Bin}(n, p)$.
  Remark: X is a random variable, x denotes a particular value of the random variable, n and p are parameters, that is, fixed real numbers. The parameter p is usually unknown and must be estimated from data.

Binomial Distribution $\mathrm{Bin}(n, p)$



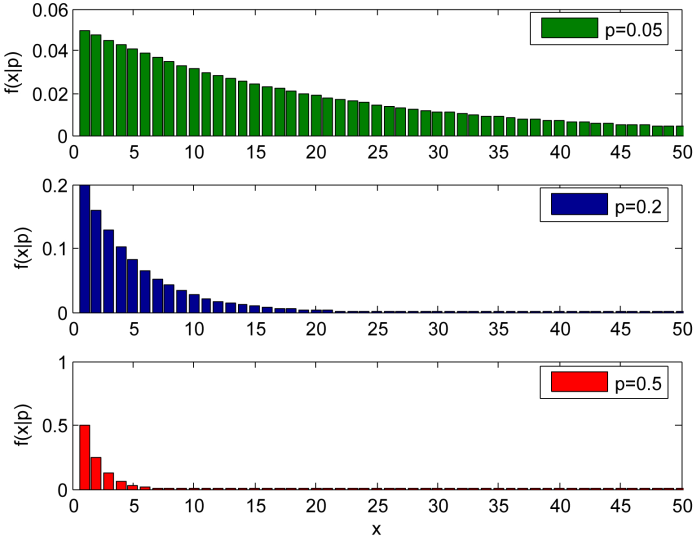
Important Examples

- **The Geometric Distribution**
  X has a geometric distribution with parameter $p \in (0, 1)$, denoted $X \sim \mathrm{Geom}(p)$, if

$$f(x|p) = p(1-p)^{x-1}, \quad x = 1, 2, 3 \ldots$$

  Think of X as the number of flips needed until the first heads when flipping a coin. Geometric distribution is used for modeling the number of trials until the first success.

Geometric Distribution $\mathrm{Geom}(p)$



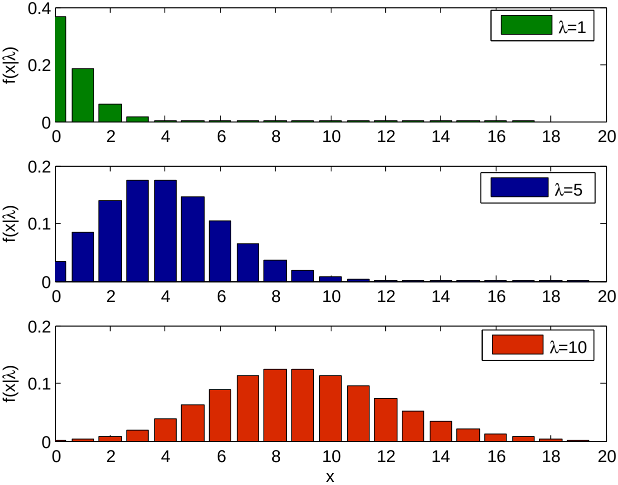
Important Examples

- **The Poisson Distribution**
  X has a Poisson distribution with parameter λ,
  denoted $X \sim \mathrm{Poisson}(\lambda)$ if

$$f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

  The Poisson distribution is often used as a model for counts of rare events like traffic accidents. $f(x|\lambda)$ expresses the probability of a given number of events x occurring in a fixed interval of time if these events occur with a known average rate λ and independently of the time since the last event.

Poisson Distribution $\text{Poisson}(\lambda)$



Summary

- A random variable is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real number $x = X(\omega)$ to each realization $\omega \in \Omega$.
- The cumulative distribution function (CDF) is defined by

$$\boxed{F_X(x) = \mathbb{P}(X \leq x)}$$

  - CDF completely determines the distribution of a random variable
  - CDF is non-decreasing, normalized, and right-continuous
- Random variable X is discrete if it takes countable many values $\{x_1, x_2, \ldots\}$.
- The probability mass function (PMF) of X is

$$\boxed{f_X(x) = \mathbb{P}(X = x)}$$

- Relationships between CDF and PMF:

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

$$f_X(x) = F_X(x) - F_X(x^-)$$